The SHARE road map

Andoulsi, I.; Blanquer, I.; Breton, V.; Dobrev, A.; Van Doosselaere, C.; Hernandez, V.; Herveg, J.; Jacq, N.; Legré, Y.; Olive, M.; Rahmouni, H.; Solomonides, T.; Stroetmann, K.; Stroetmann, V.; Wilson, P.

Published in:
Global Healthgrid

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Global Healthgrid: e-Science Meets Biomedical Informatics
T. Solomonides et al. (Eds.)
IOS Press, 2008

The SHARE Road Map[1]: Healthgrids for Biomedical Research and Healthcare

I. ANDOULSI[a], I. BLANQUER[b], V. BRETON[c], A. DOBREV[d], C. VAN DOOSSELAERE[e], V. HERNANDEZ[b], J. HERVEG[a], N. JACQ[f], Y. LEGRÉ[f], M. OLIVE[g], H. RAHMOUNI[g], T. SOLOMONIDES[g], K. STROETMANN[d], V. STROETMANN[d] and P.WILSON[e]
[a] *Centre de recherche informatique et droit, FUNDP, Namur, Belgium*
[b] *Universidad Politechnica de Valencia, Spain*
[c] *Centre National de la Recherche Scientifique, Clermont-Ferrand, France*
[d] *Empirica Gesellschaft für Kommunikations- und Technologieforschung mbH, Bonn, Germany*
[e] *European Health Management Association, Dublin, Ireland (International)*
[f] *HealthGrid, Clermont-Ferrand, France (International)*
[g] *University of the West of England, Bristol, UK*
(all affiliations at the time of the commencement of the project)

Abstract. The HealthGrid White Paper was published at the third annual conference in Oxford in 2005. Starting from the conclusions of the White Paper, the EU funded SHARE project (http://www.eu-share.org) has aimed at identifying the most important steps and significant milestones towards wide deployment and adoption of healthgrids in Europe. The project has defined a strategy to address the issues identified in the action plan for European e-Health (COM(2004).356) and has devised a roadmap for the major technological and ethical and legal developments and social and economic investments needed for successful take up of healthgrids in the next 10 years. A "beta" version of the road map underwent full review by a panel of 25 prominent European experts at a workshop in December 2007. The present document is an executive policy summary of the final draft road map. It has sought to reconcile likely conflicts between technological developments and regulatory frameworks by bringing together the project's technical road map and conceptual map of ethical and legal issues and socio-economic prospects. A key tool in this process was a collection of case studies of healthgrid applications.

Keywords. Healthgrid; Biomedical Informatics; Grid Computing – Research Challenges; Innovation and Technology Management; Ethical, Legal and Socio-Economic Issues (ELSE); Epidemiology; Innovative Medicine; Genomic & Individualised Medicine; Security; Regulatory Compliance; Collaboration

[1] The SHARE project was partially funded under the 6th Framework Programme (ICT) by the European Commission (cf. http://cordis.europa.eu/fp7/ict/).

CONTENT

1. Introduction

1.1. Summary

Grid technology has been identified as one of the key technologies to enable and support the 'European Research Area'. This new paradigm offers rapid computation, large scale data storage and flexible collaboration by harnessing together the power of a large number of commodity computers or clusters of other basic machines. The grid was devised for use in scientific fields, such as particle physics and bioinformatics, in which large volumes of data, or very rapid processing, or both, are necessary. The impact of this concept is expected to reach far beyond eScience, to eBusiness, eGovernment and eHealth. However, a major challenge is to take the technology out of the laboratory to the citizen.

The concept of grids for health was born in Europe in 2002 and has been carried forward through the HealthGrid initiative [1]. This European collaboration has edited a white paper setting out for senior decision makers the concept, benefits and opportunities offered by applying newly emerging grid technologies in a number of different applications in healthcare [2]. Starting from the conclusions of the White Paper, the EU funded SHARE project [3] aimed at identifying the important milestones towards wide deployment and adoption of healthgrids in Europe. The project has devised a strategy to address the issues identified in the action plan for a European e-Health [4] and has devised a roadmap for the major technological developments, legal and ethical barriers, and socio-economic investments needed for successful uptake of healthgrids in the next ten years.

The roadmap proposed by the SHARE project expresses certain measurable goals and objectives for the HealthGrid community, provides an analysis of the technical gaps to be bridged in order to achieve a number of staged technical objectives, explores the ethical, legal and socio-economic (ELSE) conditions of such developments, analysing the extent to which technology and its environment will need to be reconciled, and articulates a strategy for the concurrent achievement of these goals and objectives subject to realistic contextual conditions.

This roadmap has been developed from three major inputs:

a) an analysis of user requirements in a carefully triangulated set of domains through current projects and scripted use-cases;
b) a technical road map which sets out the key objectives for a viable 'knowledge healthgrid' to be achieved in a span of 10-15 years;
c) a conceptual map of ELSE conditions, constraints and requirements which must be addressed before a knowledge healthgrid can be deployed in a real healthcare setting.

The conceptual map of ethical, legal and socio-economic issues considered the regulatory challenges that any real healthgrid must meet:

- Legal challenges concerning rights to privacy and confidentiality, 'right to know' and duty of care.
- Ethical challenges concerning primary and secondary use of data whether individual or aggregated.
- Legal and ethical challenges concerning provenance and quality of information.

- Legal, ethical and economic challenges to the use of healthcare data in commercial and public research, including questions of ownership of data.
- Legal and ethical challenges in the communication of genetic information and the resultant 'lateral leakage' of information.
- Legal and ethical challenges to the communication of medical data across borders.
- Social and legal challenges concerning the formal professional competencies of different healthcare actors.
- Legal, ethical and socio-economic challenges of 'exceptional cases', such as assisted reproduction, organ donation and transplantation.

The proposed road map brings all these concerns together into one strategic plan.

1.2. What are Healthgrids?

The White Paper [2] defines the concept of a healthgrid as follows:

> *Healthgrids are grid infrastructures comprising applications, services or middleware components that deal with the specific problems arising in the processing of biomedical data. Resources in healthgrids are databases, computing power, medical expertise and even medical devices.*

A *healthgrid* is an environment in which data of medical interest can be stored and made easily available to different actors in the healthcare system, physicians, allied professions, healthcare centres, administrators and, of course, patients and citizens in general. Such an environment has to offer all appropriate guarantees in terms of data protection, respect for ethics and observance of regulations; it has to support the notion of 'duty of care' and may have to deal with 'freedom of information' issues. Working across member states, it may have to support negotiation and policy bridging.

Early grid projects, while encompassing potential applications to the life sciences, did not address the specificities of an e-infrastructure for health, such as the deployment of grid nodes in clinical centres and in healthcare administrations, the connection of individual physicians to the grid and the strict regulations ruling access to personal data. However, a community of researchers did emerge with an awareness of these issues and an interest in tackling them.

1.3. The HealthGrid Initiative

Pioneering projects in the application of grid technologies to the health area have recently been completed, and the technology to address high level requirements in a grid environment has been under development and making good progress. Because these projects had a finite lifetime and the healthgrid vision required a sustained effort over a much longer period, and besides because there was an obvious need for these projects to cross-fertilise, the 'HealthGrid initiative', represented by the HealthGrid association (http://www.healthgrid.org), was initiated to bring the necessary long-term continuity. Its goal is to encourage and support collaboration between autonomous projects in such a way as to ensure that requirements really are met and that the wheel, so to speak, is not re-invented repeatedly at the expense of other necessary work.

The HealthGrid community identified a number of objectives [5]:

- Identification of potential business models for medical grid applications.
- Feedback to the grid development community on the requirements of the pilot applications deployed by the European projects.
- Development of a systematic picture of the broad and specific requirements of physicians and other health workers when interacting with grid applications.
- Dialogue with clinicians and those involved in medical research and grid development to determine potential pilots.
- Interaction with clinicians and researchers to gain feedback from the pilots.
- Interaction with all relevant parties concerning legal and ethical issues identified by the pilots.
- Dissemination to the wider biomedical community on the outcome of the pilots.
- Interaction and exchange of results with similar groups worldwide.
- The formulation and specification of potential new applications in conjunction with the end user communities.

Apart from research, where the value of grid computing is well established, a healthgrid may be deployed to support the full range of healthcare activities, from screening through diagnosis, treatment planning to epidemiology and public health. For example, anticipating that population trends, air pollution and global warming may lead, through extremes of heat, to increased risks for the elderly, we may deploy a monitoring service to track conditions and medical episodes in hot summers.

The results of several major studies of the interface between bioinformatics and medical informatics have been published with a remarkable promise of synergy between the two disciplines, leading to what had already begun to be referred to as 'personalised medicine' [6–8]. From the point of view of HealthGrid, this made clear the need to unify the field and to put its various elements in perspective: how would they – improved evidence bases, imaging, genetic information, pharmacology, epidemiology – fit together, what was their relative importance in the unfolding programme of work?

Given the source of the concept of grid in the physical sciences, many requirements arising out of the biomedical and healthcare fields were not a central concern to the grid development community. Indeed, even today, when these requirements have been fed through to the middleware services community, they are not always or necessarily a priority for the developers. Thus HealthGrid has been actively involved in the definition of requirements relevant to the development and deployment of grids for health and was among the first to identify the need for a specialist middleware layer, between the generic grid infrastructure and middleware and the biomedical or health applications.

Among data related requirements, the need for suitable access to biological and medical image data arose in several early projects, but for the most part these are present in other fields of application also. Looking to security requirements, most of these are special to the medical field: anonymous or private login to public and private databases; guaranteed privacy, including anonymization, pseudonymization and encryption as necessary; legal requirements, especially in relation to data protection, and dynamic negotiation of security and trust policies while applications remain live. Most administrative requirements are common to medicine and eScience, although the flexibility of 'virtual grids', i.e. the ability to define sub-grids with restrictions on data storage and data access and also on computing power, is more obviously required in healthcare.

Medical applications also require access to small data subsets, like image slices and model geometry. At the (batch) job level, medical applications need an understanding of job failure and means to retrieve the situation.

1.4. The SHARE Project: From White Paper to Road Map

In the White Paper, the HealthGrid community expressed its commitment to engage with and support modern trends in medical practice, especially 'evidence-based medicine' as an integrative principle, to be applied across the dimensions of individual through to public health, diagnosis through treatment to prevention, from molecules through cells, tissues and organs to individuals and populations. In order to do this, it had to address the question how to collect, organise, and distribute the 'evidence'; this might be 'gold standard' evidence, i.e. peer reviewed knowledge from published research, or it might be more tentative, yet to be confirmed knowledge from practice, and, in addition, would entail knowledge of the individual patient as a whole person. The community also had to address the issues of law, regulation and ethics, and issues about crossing legal and cultural boundaries, finding ways to express these in terms that translate to technology – security, trust, encryption, pseudonymisation. Then it had to consider how the services of the healthgrid middleware would satisfy these requirements; and, if it was to succeed in the real world, how to make the business case for healthgrid to hard-pressed health services across Europe while they are struggling with their own modernisation programmes.

The vision of health that informs the thinking of the White Paper and the work of HealthGrid since its publication has been defined in the 'Action Plan for a European e-Health Area' [4] as follows:

> *"... the application of information and communications technologies across the whole range of functions that affect the health sector. e-Health tools or 'solutions' include products, systems and services that go beyond simply Internet-based applications. They include tools for both health authorities and professionals as well as personalised health systems for patients and citizens. Examples include health information networks, electronic health records, telemedicine services, personal wearable and portable communicable systems, health portals, and many other information and communication technology-based tools assisting prevention, diagnosis, treatment, health monitoring, and lifestyle management."*

In the light of the White Paper and its impact, the EC funded a 'specific support action' project, SHARE, to explore exactly what it would mean to realise the vision of the White Paper, investigate the issues that arise and define a roadmap for research and technology which would lead to wide deployment and adoption of healthgrids in the next ten years.

2. The Benefits of Healthgrids

2.1. For the Healthcare Professional/Biomedical Researcher

Healthcare systems both in developed and in developing countries face major economic and capacity challenges to maintain quality of care in the face of the growing demands of ageing populations and the increasingly sophisticated treatments available. Add to this the desire to improve access to new care methods, and the challenge of delivering

care becomes significant. In an attempt to meet these demands, health systems have increasingly looked at deploying information technology to scale resources, to reduce queues, to avoid errors and to provide modern treatments into remote communities, for example.

From the individualized care point of view, in order for clinicians to make the best diagnosis and decide on treatment all the relevant health information of the patient needs to be available and transparently accessible to them regardless of the location where it is stored. Moreover, computer-aided tools are now essential for interpreting patient-specific data in order to determine the most suitable therapy from the diagnosis [9].

To store and process medical images, genetic information and other patient data, a large amount of computing power is needed. Large computing resources are also needed for keeping statistics of patient records, for knowledge extraction using data mining, and for the simulation of organisms and diseases using complex biomedical models. Grid technology has undoubtedly much to offer medical professionals, as illustrated by the following examples [11].

However, the modernisation process faces significant challenges:

- Connecting and understanding patient records across organisation structures and even national borders.
- Ensuring that information is secured and those accessing it are authorised and authenticated.
- Discovering trustworthy sources of information for comparison.
- Handling a huge volume of data, especially that involved in genetic medicine for instance.
- Applying traditional information networks and technology into healthcare.

The delivery of medical information and certain services through the internet is familiar. In healthgrid computing, we seek an extension of the concept to consider how to provide large scale services to the user on demand. Some examples will serve to illustrate:

i. Consider a radiologist who needs to manipulate an image: we want to provide a set of services, some of which may require heavy processing, making them available on her desktop 'transparently', as if they were programs simply running on her computer.
ii. Consider a public health service which monitors certain infectious diseases and has to trigger an alert in case of a suspected epidemic. The identification of unusual patterns would in many cases be the critical step to halting the problem.
iii. Consider a surgical simulation prior to maxillofacial surgery, to determine how the patient's face may appear after one manoeuvre versus another, the presence of sufficient tissue to allow the operation or to demand transplantation, and even to involve the patient in the decision.
iv. Consider a 'neglected disease' like malaria. Malaria is neglected by the pharmaceutical industry because there is no prospect of profit in it. Relatively little progress has been made towards the eradication of this well understood disease, notwithstanding substantial investments of public funds in research projects. *In silico* lead generation may possibly be coupled with investment in

plant by the poorer nations that suffer from it to lead to a locally sustainable solution.

v. Consider the possibility of linking genomic information to imaging in diseases like juvenile idiopathic arthritis. The genome will indicate susceptibility long before the disease is expressed, but equally, signs picked up from imaging may obviate the need for genetic screening, thus avoiding some of the most acute problems associated with it.

vi. Consider more abstractly the nature of evidence-based practice, the volume of scientific literature that provides the evidence base and the accumulation of evidence from practice that occurs as a matter of routine healthcare. How can these be integrated? How can they be used without violating any ethical restrictions on use of data, confidentiality, privacy, security? How can they be shared without violating any data protection laws?

However, there are problems even among these optimistic scenarios. Standards have not stabilised in the grid world, so data exchanges will present problems straight away. Codes and coding languages are also still not universally adopted, while the application provider will wish to protect investments in software licence rights.

2.2. For the Technologist

The SHARE roadmap for the adoption of healthgrids constitutes a critical analysis of the status of grid and other supporting technologies for the advance on the integration and processing of large scale eHealth and biomedical data.

From the technologist's point of view, this document outlines the deficiencies, gaps and promising technological research lines that are necessary for achieving a reasonable degree of maturity in healthgrids. Therefore, it can be seen as a list of opportunities for collaborations, new working lines and technology transfer.

In this sense, four technological areas are considered:

- Computing challenges. Issues related to the reliability, quality of service, lightweight middleware and compatibility on health networks require specific actions which are outlined.
- Data grid challenges. Issues related to data federation, effective update of databases, scalability and privacy management are considered and analysed.
- Collaboration grid challenges. Issues related to workflow definition, threading processes, 'playing' with data, adjusting images, consulting colleagues, comparing and contrasting, have been analysed at greater length in eScience projects than in healthgrids.
- Knowledge challenges. The evolution to future knowledge services through the semantic integration of services, semantic data analysis and federation are medium and long term issues which should be started now through basic research.

These technological challenges are summarised in several milestones that are described in the final section of the roadmap.

2.3. Socio-Economic Benefits

Modern healthcare services are expected to be available around the clock, seven days a week, so that systems with pervasive access and near-absolute fault tolerance are indispensable. However, it is difficult for these applications to run non-stop with a high quality of service. Grids could help by providing a platform of collaboration, allowing the linking centres which co-operate to achieve better continuity and quality of service. Medical staff will then be able to share experience, knowledge and 'second opinion' with other internal and external staff. The distributed architecture of grids with the availability of high-bandwidth networks responds well to the requirements of healthcare provision. There are also optimistic stakeholders' views towards medical research, healthcare and computing capabilities combined to better satisfy the patient [10].

Healthgrids promise many benefits to mobile patients as well as citizens. It could help a travelling individual to receive the right treatment in an emergency situation, thanks to the ability of the grid to facilitate communication between the local hospital of the patient and the admitting hospital abroad in order to exchange necessary heath related information.

In addition, healthgrids enable the mobility of a patient within EU states and allow them to receive medical treatment in a country of their choice. This could help solve problems of long waiting lists in states with busy hospitals and lack of medical staff. In economic terms, the grid could provide an optimal solution for healthcare. It allows a better use of resources and maintenance of tasks, an improved global IT organisation, scalable costs, and a large and consolidated IT business within the healthcare organisation [11].

Heath tourism is a growing concept which can enrich the economy of countries where modern medical treatments (plastic surgery, dental surgery, reproductive medicine, laser surgery for vision correction, etc.) are evolving and having higher success rates than others. This domain can benefit from healthgrid technology as it facilitates the exchange of patient heath records and the communication between foreign hospitals and heath insurance companies to facilitate the referral and payment process.

Transferring medical images for the purpose of a second opinion to another hospital requires high bandwidth connections between hospitals. Healthgrid technology can provide automated workflows that could be considered a better alternative to manual workflows, such as agreement over the phone and fax transmission of data. These manual workflows are still used by clinicians at present, but are labour-intensive and can cause errors [10].

3. Use Cases, User Requirements and Challenges

3.1. Grid Paradigms

Grids are often differentiated into computational, data and collaboration grids. The ideal grid, envisaged as a servant of a new paradigm of scientific research called 'e-science', must provide transparent processing power, storage capacity and communication channels for scientists who may from time to time join the grid, do some work and then leave, so that the alliances they form in their scientific endeavours might be described as 'virtual organisations' or VOs for short. Different sciences have different needs, and the grid concept has become differentiated: particle physics generates enor-

mous amounts of data which must be quickly stored, but not necessarily instantly processed; on the other hand, data in bioinformatics is not large by comparison – it is, of course, in plain terms, large – but requires intensive processing. In extending the application of grid computing to e-health, another feature becomes pre-eminently necessary: that of collaboration.

3.2. Computing Grids: An Example from Innovative Medicine

Drug discovery is the long term, multi-stage and high cost process by which drugs are discovered and/or designed. The drug discovery goal is to find new molecules that bind with specific macromolecules known to play a key role in a disease process, in a manner that changes their function, either to increase resistance to or to reduce the virulence of some pathogen. Reducing the research time in the discovery stage and having enhanced information about substances affecting the selected target ('leads') are key priorities for pharmaceutical companies worldwide.

In silico drug discovery, including analysis of the gene expression data, target function prediction and target three-dimensional (3D) structure prediction, is one of the most promising strategies to speed up the drug discovery process, avoiding time consuming and costly *in vitro* and *in vivo* tests. *In silico* drug discovery contributes to increasing biological system knowledge, to managing data in a collaboration space, to speeding up analysis and consequently improving the success rate compared with the traditional "wet" approach. The efficiency gains of such an integrated knowledge system could result in 35% cost savings, or about US$300 million, and 15% time reduction, or two years of development time per drug.

In silico drug discovery requires advances in **data integration** (including data format standardisation, dataflow definition in a distributed system, services for data and meta-data registration, and development and sharing of ontologies and knowledge representations), **workflow enactment** to ease data management and data mining and to assist the scientist and the decision-maker in organising their work in a flexible manner, **access to computing and data resources** (computing 1 million docking probabilities or modelling 1000 compounds on one target protein requires a few TFlops for one day), and **collaboration between public and private partners**, involving the concrete sharing of data and knowledge, software and workflow, and infrastructures such as computing, storage and networks. Security and the effective protection of intellectual property and sensitive information are key challenges for pharmaceutical industries, but also for academic institutes in most cases.

3.2.1. Challenges and Requirements

The European Federation of Pharmaceutical Industries and Associations (EFPIA) identified pre-competitive barriers to innovation. The objective for the future would be to identify as soon as possible in the pre-clinical phase the reasons for lack of efficacy, despite promising pre-clinical data, and the potential for adverse drug reactions and pre-clinical toxicity. The identified key bottlenecks in the R&D process are:

- predictive pharmacology at the discovery research stage;
- predictive toxicology at the preclinical development stage;
- identification of biomarkers at the translational medicine stage;

- patient recruitment and validation of biomarkers at the clinical development stage;
- risk assessment with regulatory authorities at the pharmacovigilance stage.

The knowledge management area is identified as key to leveraging the potential of new technologies such as genomics and proteomics and to analyse the huge quantity and diversity of information in an integrated way, such as the capture, analysis and interpretation of knowledge generated for one potential drug candidate from discovery, non-clinical and clinical development all the way to lifecycle management. One of the major bottlenecks is the lack of availability of databases across R&D that might facilitate data integration.

3.3. Data Grids: An Example from Epidemiology

One relevant example of health applications for data grids is epidemiology. The epidemiology use case is defined as a system able to link the information from distributed and heterogeneous databases, identify patients, complete episodes and automatically improving quality without interrupting clinical practice. With this data, complex epidemiological models are fed and simulated producing aggregated prospective results, in a reliable way. This use case is representative of different applications and systems including ontological information systems, infectious surveillance networks, pharmaepidemiology analysis of efficiency and cost, and study of propagation models for diseases.

The main users (from the highest-concept level to lowest) are public health authorities, epidemiologists and pharmaceutical companies. The data is normally owned by clinical care (both public and private). In order to achieve this use case the user needs to go through different steps including automatic data gathering, data quality improvement, processing of the data, and presentation of results.

3.3.1. Challenges and Requirements

The main challenges to this use case include the general problem of access to distributed, critically sensitive and heterogeneous data, resulting in costly computing processes. Patient-centric analyses normally deal with smaller amounts of data and require a pre-existing knowledge of models of healthy and diseased organs or tissues. Population-level analyses normally deal with the integration of larger, poorer-quality data. Semantics are especially relevant in those approaches.

The research requirements for the epidemiology use case can be summarised as follows:

- Effective semantic annotation of data. Data is poorly coded and interoperability of coding is not trivial. Extracting knowledge from medical data, however, is a main objective.
- Effective integration of distributed and heterogeneous data. Integrating distributed resources requires exchange protocols, secure mechanisms, patient de- and re-identification, and automatic data analysis services.
- Availability of efficient infrastructures and usage policies. Applications will require sufficient resources and a reliable infrastructure to work on under a clear Quality of Service (QoS) promise.

- Robust security mechanisms insuring that the different components of the infrastructure behave in way that conforms to the European norms controlling the handling of personal data.
- User-friendliness of applications and services. The tools should be available through protocols and interfaces similar to those used in the users' normal research. Not only must the applications be as compliant as possible with current systems and interfaces, but so must the technologies.
- Ensuring that the research is done in a secure, ethically and legally compliant framework.
- Reliability, scalability and pervasiveness. All the previous services must be robust and should be scalable without reducing performance.

3.4. Collaboration Grids: AN Example from Virtual Physiological Human (VPH)

The EuroPhysiome initiative has led to the concept of Virtual Physiological Human (VPH), indicating a methodological and technological framework that once established will enable the investigation of the human body as a single complex system. VPH will provide a framework within which observations made in laboratories, hospitals, and in the field all over the world can be collected, catalogued, organised, shared and combined in a variety of ways. It should also allow experts to collaboratively analyse observations and develop systemic hypotheses that involve the knowledge of multiple scientific disciplines, and to interconnect predictive models defined at different scale, with different methods, and with different levels of detail, into systemic networks that provide concretisation to those verifiable systemic hypotheses.

3.4.1. Challenges and Requirements

With respect to grid computing, the following challenges and requirements have been identified.

Knowledge management. The accumulative nature of VPH requires that everything is organised within solid knowledge management models. This should make it possible to keep even very large information spaces organised and usable.

Fostering grid adoption. To foster grid adoption in the VPH community, it is highly recommended to identify a few VPH CPU intensive applications which could benefit immediately from deployment on existing grid infrastructures like EGEE or DEISA. The deployment of these applications will allow the identification of missing services on the existing infrastructures, and will raise the awareness of grids in the VPH community.

Access to resources. Researchers require access to all available resources in a uniform way, from those provided by their own department to specialised HPC resources. Access to these should be as seamless as possible, with simulations at different scales being automatically migrated and appropriate resources being used as required.

User friendly interfaces. Current grid portals require the user to specify parameters such as memory to be allocated at execution time; this would not be appropriate for VPH users.

Grid usage models. The nature of VPH simulations means that processing is time critical, and current models of HPC use would not be appropriate. Instead, models which permit a large number of grid nodes to be used for a relatively short time ('burst mode') with little or no waiting time should be established. Resource co-reservation

will also be required, particularly where multiscale simulations that run over multiple sites are concerned.

Shared storage for large data/model repositories. The imaging datasets concerned can be several hundred megabytes in size, but after pre-processing to generate a predictive model, the modelling and simulation data can be as much as a hundred gigabytes in size. With potentially thousands of these, there is a clear need for multi-terabyte storage facilities, connected to distributed HPC resources via high speed networks. These must incorporate the required security and confidentiality measures.

Methods to solve multiple predicted models, in a coupled way. As the coupling of predictive models at different scales is central to VPH's description of human physiology, coupled methods to solve multiple models will be required. This is considerably complicated by the fact that the models concerned may be very different both in conceptual nature and mathematical nature. Even relatively simple VPH problems can be considerably complicated by variations between individual subjects and treatment procedures.

Direct prediction from medical images. Methods for transforming a medical imaging dataset into a subject specific predictive model that do not require the costly pre-processing phase are being developed, such as the Boltzmann Lattice in haemodynamics and voxel meshes for hard tissue simulations. However, these are enormously computationally intensive, requiring fifty or more teraflops of computational power to solve in less than a day.

3.5. Knowledge Grids: AN Example from General Healthcare

For any given domain, a distinction is often drawn between declarative knowledge ('know what') and procedural or operational knowledge ('know how'). In the domain of healthcare, both kinds occur. What is often referred to as 'the scientific basis' of medicine, that which must furnish the evidence in so-called 'evidence-based practice', is present in research publications to which different standards of credibility are attached. For example, research results based on a randomized, double-blind, controlled clinical trial are held to be the gold standard, provided they were also submitted to adequate peer review. Evidence based on one physician's own practice, although not negligible, would be considerably less reliable. On the other hand, 'best ways' of treating patients – in a particular context – are often described in integrated care pathways (ICPs). It is not unreasonable to claim that declarative knowledge in medicine tends to be disseminated through peer-reviewed publication and operational knowledge through such things as guidelines and care pathways. In a healthgrid environment, these are brought together for the better treatment of patients and at the same time to improve research; indeed, the interplay between healthcare and research, e.g. through appropriately controlled 'secondary use', would be an important element in a full healthgrid environment.

Chronic obstructive pulmonary disease (COPD) refers to an airway obstruction caused by chronic inflammation. It is usually progressive, not fully reversible, and often occurs as a result of smoking but other factors such as air pollution can also contribute to the development of COPD. In the UK almost 900,000 people have been diagnosed with the disease, and the true number of people suffering from the condition is estimated to be around 1.5 million.

According to NHS guidance, the management of the disease should be tailored to the individual, with adjustments being made based on responses to treatment. The

guidance includes a large number of drugs, including some off-label drugs such as Beclometasone, Fluticasone and Budesonide.

In the UK, the National Institute for Health and Clinical Excellence (NICE) has issued national guidelines for the treatment of the disease, but these are frequently modified to account for local variations and priorities. As a result, the procedure for assessing and treating the disease will vary even within a single country, let alone between countries. The evidence on which this guidance is based comes from a variety of sources, such as national studies by NICE and systematic reviews with an international scope.

Two main concerns exist for general healthcare; supporting the travelling patient, such as migrating elderly populations, and enabling decision support systems that can account for local variations in best practice and clinical evidence.

3.5.1. Challenges and Requirements

Evidence from national studies, such as the aforementioned NICE study, may not be available to a doctor from a different locale. In order to continue treating the patient concerned, the doctor (or decision support system) must be aware of the evidence and guideline/pathway that informed the plan of care for that patient, and any deviations from that plan that have occurred to date. This may not a trivial matter of simply retaining a link to the relevant material, as there may be language barriers, and local reasons why the guidance followed in one country would not be appropriate in another.

The guidance also mentions drugs that are not certified for the treatment of COPD (off-label) despite this evidence coming from high quality systematic reviews. Different drugs will be certified for the treatment of the condition in different countries, complicating the process of following a single guideline or pathway regardless of travel. In fact, the patient concerned may be travelling for the express purpose of receiving different or less costly treatment in another country.

Prior history of exacerbations and smoking are essential for properly treating the disease, and therefore the doctor concerned must be able to access, comprehend and update the patient's record. This requires standardisation of electronic health records (EHRs) and electronic integrated care pathways (eICPs). When it comes to decision support, a standard interface format, such as the proposed HL7 vMR [15], will also be a necessity.

4. Technical, Ethical, Legal and Socio-Economic Issues

4.1. Technical Issues

4.1.1. Standardisation Issues

Standards are absolutely necessary for the deployment of services which integrate data in bioinformatics and medical informatics, but are also vital for data coming from different medical disciplines and even data coming from different countries in Europe. These standards are needed for building data models, producing ontologies and for the development of knowledge management services. The adoption of standards for the exchange of biological and medical information is still limited to a few specific fields. Moreover, they need to be compatible with grid standards so as to allow their implementation on healthgrids.

4.1.2. Communication Issues

Lack of information about grids and grid technologies is frequently identified as one of the key reasons why there has been very little interest in them from the field of medical research. It is essential that all relevant actors to be kept well informed by the HealthGrid community of the potential benefits of the technology to them. Success stories demonstrating the impact of grids for medical research will be vital for convincing medical researchers of these benefits. As a result, there is a need for a suitable demonstration environment, offering very easy access to the grid for non experts and providing services that will help convince the medical research community. On this dissemination environment, dedicated efforts to promote the technology can then be developed.

4.1.3. Security Issues

Deployment of a data grid for medical research will only be possible when the middleware can provide all the necessary guarantees in terms of management of personal data. We perceive the specific technical requirements related to the handling of medical data on the grid to be as follows.

- Manipulation of personal data on the grid must obey strict regulations. These regulations vary between European member states.
- Services for the anonymization and pseudonymization of medical data must be provided.
- Medical data is the property of the patient. A mechanism must be set up to allow individuals to access their data on the grid.
- For healthcare purposes, the authentication of healthcare professionals on the grid cannot be handled by requesting all of them to get a grid certificate. A mechanism must be set up so that professional cards can be used to provide authentication on the grid.

4.2. Legal Issues

4.2.1. Data Protection

The ethical principle of autonomy is legally underpinned by the duty of data protection. This is taken very seriously at EU level; as well as the inclusion of privacy protection in the European charter of fundamental rights, the EU has developed the robust directive on data protection to promote privacy.

However, the current legislation is not adequate to support most of the longer running research initiatives around which healthgrids are based [12]. As the current EU level legislation stands, member states can enact specific legislation covering specific tools such as healthgrids in order to exempt scientists and medical practitioners using healthgrids from some of the more onerous duties of the directive.

No member state has addressed legislation to this particular issue and so healthgrids are burdened with onerous data protection requirements which could deter scientists from adopting healthgrid technology and using its enhanced computational and data acquisition power.

4.2.2. Liability

EU level legislation on liability for goods and services is reasonably well developed, but does not in its present form lend itself well to the healthgrid domain. One of the reasons for this is, of course, that health services are organised at national or regional level and that the European Union has no legal competence to draw up legislation which states specifically how a health service should be organised. However, the EU does have a range of legislation designed to protect citizens from harm resulting from goods offered on the market [13]. Steps could be taken using guidelines, or even specific legislation, to address distributed computing services such as healthgrids that would seem at present to be only marginally covered by the existing rules. Accordingly, it is important that the existing European framework of general product safety be re-examined to consider its applicability to distributed networks such as healthgrids.

4.2.3. Intellectual Property

In the EU, legislation dictates that the owner of copyrighted software running a healthgrid has exclusive rights to reproduce his work, produce derivative works, distribute copies to the public, and perform and/or display the work publicly. Under these circumstances any natural or legal person would have to pay to use the computer programs that constitute one of the most important components of healthgrids.

An open standards approach to software co-development could help in the development and implementation of healthgrids. The open source licensing model actually uses copyright and contract principles to retain control of the work while enabling its use effectively for free, and could thus encourage use and development.

4.3. Organisational, Social and Cultural Issues in the Use of Healthgrids

Both at the individual and the societal level, issues like universality of availability of full healthcare services to all citizens, equal access to healthcare, and equally high quality of services rendered are key issues [14]. Geographical factors relate mainly to equal access to quality care independent of location. ICT-based systems pose new problems like access to EHRs by insurance companies or employers, and even police and prosecutors. The opinions and attitudes of patient and citizen associations and lobbying groups, often magnified by the media, could have a strong impact through public (policy) discussions of these topics on the implementation and diffusion of healthgrids.

The organisational level is always complex. Perspectives, confirmed by two recent research studies, include:

- Changing care pathways that need new information, skills, knowledge and processes for healthcare providers.
- Changing roles of healthcare professionals, teams and healthcare organisations.
- The transfer of roles between healthcare professionals, teams and healthcare organizations.
- Increased collaborative working and exchange of information between providers.
- New relationships between citizens and healthcare professionals and organisations.
- New strategic partnerships for third party financers and healthcare providers.

Cultural issues are a key factor in health services, including the great diversity of attitudes, behaviour and knowledge exchange among professional and non-professional staff involved in healthcare, and the impact this has on the quality, efficiency and organisation of services. Education and training, professional standards and bodies, rules and regulations, attitudes and behaviour all have an influence here.

5. Recommendations and Roadmaps

Objectives have been formulated in terms of milestones according to a number of key criteria:

- Is the proposed healthgrid essentially a computational grid, a data grid or a collaboration grid? Could it potentially develop into a knowledge grid?
- Is the necessary development to achieve any given stage likely to be delivered by generic grid research or is it particular to healthgrids?
- Is some prerequisite standard or other agreed framework necessary for the achievement of any particular milestone?

5.1. Health Research Challenges from User CommunitIES' Requirements

The analysis of the user community requirements documented in the previous section show very clear patterns:

- Knowledge management is what researchers need. Computing and data storage resources are not sufficient although it is expected they can be accessed in a transparent and ubiquitous way;
- Although the existing grid infrastructures do not provide all the services needed by the user communities, they already permit a number of tasks of scientific relevance. As a consequence, deployment of scientific applications should be started as soon as possible in order to foster grid adoption and to clearly identify the existing gaps;
- The technological complexity must be hidden from users. Grids are perceived as potential infrastructures in so much as their use does not require adaptation or acquisition of skills;
- The communities examined expressed the need for developing the technology for distributed data management, and while the usage of grids for distributed computing is perceived as available it is still very complex.

In the rest of this section, we have attempted to translate the requirements of three communities (epidemiology, innovative medicine and VPH) into a number of health research challenges and deployment milestones.

- The health research challenges are technical issues which need to be addressed in order for grids to offer services needed by the health communities.
- The health deployment milestones are health applications that should be deployed on grids in order to demonstrate their relevance, to identify existing limitations and to quantify the progresses made.

The research challenges have been classified according to their relevance to computing, data, collaboration and knowledge grids. We have also identified several that

Table 1. Health Research challenges for a computing grid

Research challenge name	Description of the health research challenges
RCCG1	Automatic migration of simulations between different scales and platforms
RCCG2	Capacity to access grid resources on demand, without previous agreement or request. European grid infrastructures should be freely accessible to European projects
RCCG3	Capacity to submit jobs to cluster and supercomputer grids in a transparent way. Easy transfer of tasks between grid infrastructures
RCCG4	Transparent access. The users should ignore they are using one grid or the other
RCCG5	User friendly access. Lower barrier to adoption.
RCCG6	Real fault-tolerant scheduling systems
RCCG7	Grid middleware that can be installed in health environments seamlessly and without requiring exhaustive maintenance and administration.
RCCG8	Services in the infrastructures to define exploitation models and guarantee a Quality of Service. Need to consolidate the booking of resources in advance and to guarantee a pre-negotiated Quality of Service.
RCCG9	Scalable job scheduling system
RCCG10	Integration of resources with low latencies and high performance.

are not specific to grids but which are needed for the deployment of knowledge grids, such as the definition of agreed standards and ontologies within research communities.

5.1.1. Health Research Challenges for Computing Grids

Table 1 lists the research challenges identified from the requirements expressed by the research communities for computing grids (RCCG). They focus mainly on user friendliness, interoperability, quality of service and on demand access:

- User friendliness (RCCG5) is needed in order for the communities to use the grids without having to learn complex procedures. To make the grid user friendly, its operating system must be fault tolerant (RCCG6). The complexity should be hidden to the point the use of grids become transparent (RCCG4, RCCG3).
- The need to access resources on clusters and supercomputers raises the need for interoperability between grid infrastructures (RCCG3). The transfer of jobs between infrastructures should also be made transparent to the user to ease his work (RCCG1).
- The quality of service is particularly critical for biomedical applications in relation to healthcare (RCCG8). This includes the need for a scalable job scheduling system (RCCG9), the availability of a robust middleware easy to install in health environments (RCCG7) as well as resources with low latencies and high performance (RCCG10).
- On demand access to the resources (RCCG2) raises technical and political as well as financial issues as to who pays for operating the infrastructures.

The four key words we will keep to characterise the research challenges for computing grids are user friendliness, interoperability of infrastructures, quality of service and on demand access.

Dependencies for these challenges, grouped by key words, can be seen in Fig. 1.

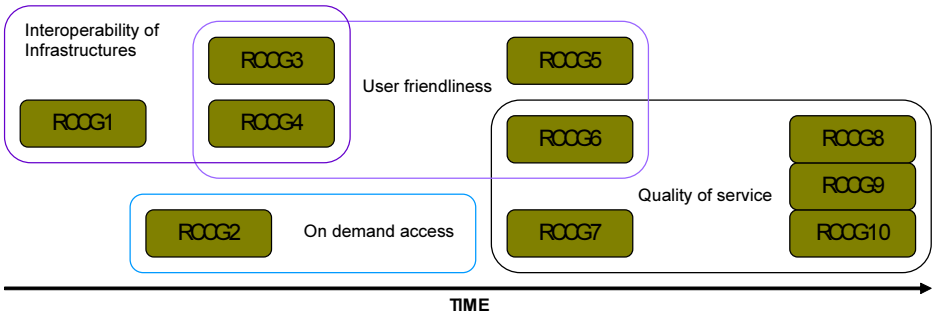


Figure 1. Dependencies for computing grid research challenges.

Table 2. Health Research challenges for a data grid

Research challenge name	Description of the health research challenges
RCDG1	Scalable data cataloguing and data transfer.
RCDG2	Storage services designed to ease the upload and download of large binary objects
RCDG3	Develop enhanced standards for data protection in a web services environment
RCDG4	Grid middleware that can be installed in health environments seamlessly and without requiring exhaustive maintenance and administration.
RCDG5	Services in the infrastructures to define exploitation models and guarantee a Quality of Service. Need to consolidate the booking of resources in advance and to guarantee a pre-negotiated Quality of Service.
RCDG6	Data architectures and tools that implement private data dissociation, pseudo-anonymisation and encryption, and that are able to fulfil the legal requirements in the matter of data management.
RCDG7	Distributed data models and repositories adapted to the multiscale nature of the data needed and generated by the health community

Although largely arranged by level of complexity, certain milestones are prerequisites for others. For example, for true transparent access to multiple grids and infrastructures (RCCG4) and similarly the transfer of tasks between infrastructures (RCCG3), it will first be necessary to enable on demand access to grid infrastructures without prior agreement (RCCG3).

It should be noted that work towards achieving these milestones is expected to be done in parallel. Although this ideal timeline reflects dependencies (to some extent), it is also the case that the demand for various developments arises from different quarters with plans and development programmes working towards their achievement at different stages of progress. In any case, we observe that computing grids are at a more advanced stage in their development in general, so that work in progress here may fairly be expected to support and facilitate progress in data grids and, as they emerge, knowledge grids.

5.1.2. Health Research Challenges for Data Grids

Table 2 presents the research challenges for a data grid (RCDG). Some of these challenges seem to be common to computing grids like the need for quality of service (RCDG5), including the availability of a robust middleware easy to install in health environments (RCDG4). But these challenges require different skills and content.

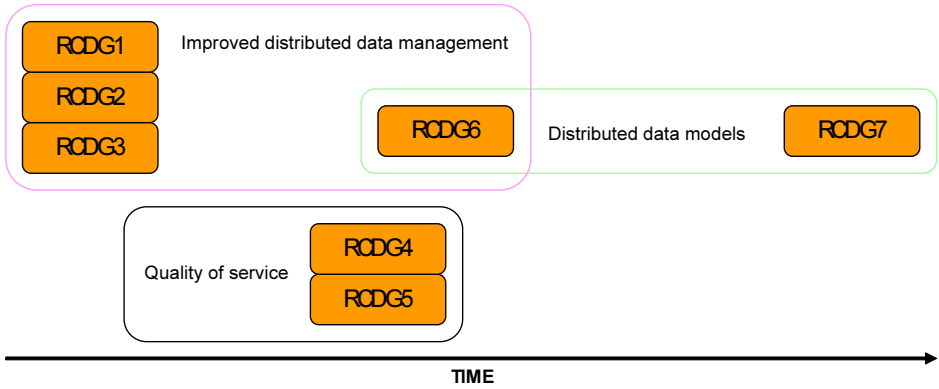


Figure 2. Dependencies for data grid research challenges.

Some challenges are related to basic data management services which are still to be developed such as scalable data cataloguing and data transfer (RCDG1) as well as upload and download of large binary objects (RCDG2). Further developments include services to provide security in the management of the medical data (RCDG6) related to the adoption of standards (RCDG3).

The need for distributed data models (RCDG6, RCDG7) is also expressed.

The key words we will keep to characterise the research challenges for data grids are improved distributed data management, quality of service and distributed data models.

The dependency diagram for these milestones, grouped by key words, can be seen in Fig. 2.

It should be noted that quality of service (QoS), a key word for computing grids, is also an important area for data grids. The milestones RCCG7/RCDG4 and RCCG8/RCDG5 respectively are similar, although there will be differences in the specific requirements for QoS between computing and data grids.

Naturally, there is a significant emphasis on the handling of data. Most questions will have already occurred in some guise or other in the field of distributed databases, but they reappear here with force in view of the autonomy of nodes within virtual organisations and especially the critical control that (non-virtual) organisations in the healthcare and biomedical domains must exercise over their data.

Developments in computing grids are anticipated to support the evolution of data grids, although there is no simple correspondence between the different concerns and drivers in the two paradigms. Indeed, it is important to observe that the principal concern of data grids, the management of its transparently distributed data, may be addressed in parallel with the majority of issues in computing grids. This is happening in several quarters in some cases independently of computing grid research and elsewhere in relation to it. Health-related projects dealing with imaging in particular, such as the EPSRC-funded Integrative Biology project [14] and the EC-funded Health-e-Child project [15], have features common to both computing and data grids. These require significant data management facilities for distributed, possibly heterogeneous image data, associated annotations and metadata, but also require computational resources for biomedical modelling and simulations.

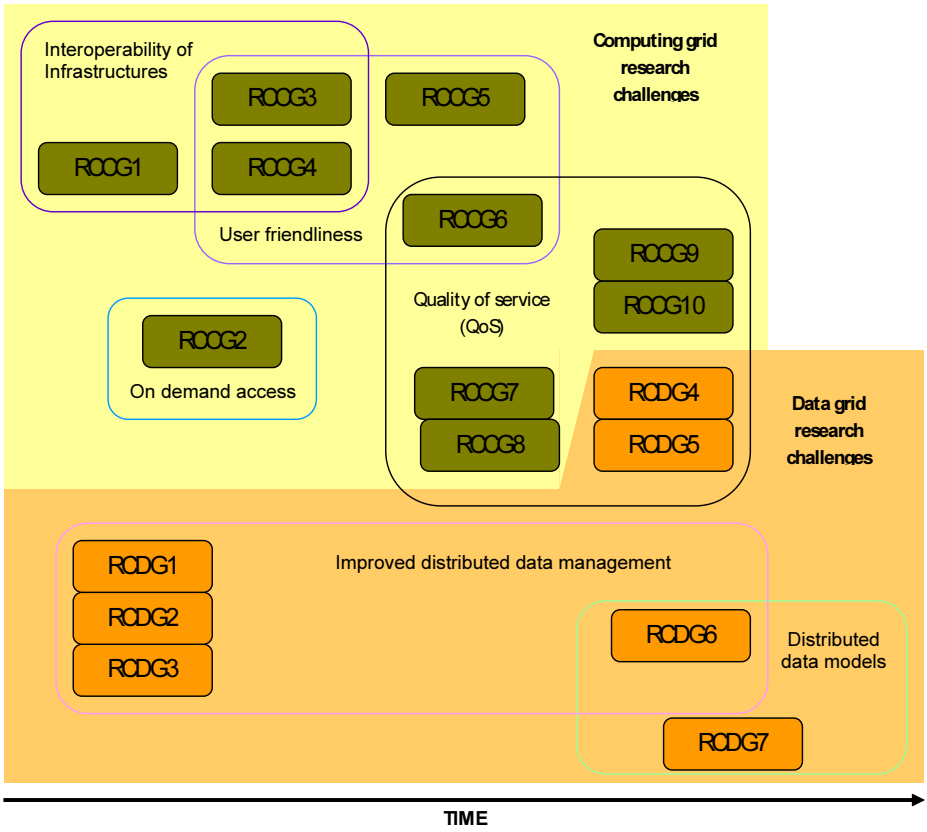


Figure 3. Combined dependency diagram for computing and data grid research challenges.

As noted previously, developments in computing grids will support some of the work still necessary in the domain of data grids. Fig. 3 illustrates the overlap between them.

5.1.3. Health Research Challenges for Collaboration Grids

Table 3 below catalogues the principal research challenges for collaboration grids, i.e. for grid services to support collaboration (RCLG). Biomedical research and healthcare are often highly cooperative, multidisciplinary activities, underpinned by informal as well as formal networks. While in some healthgrid projects collaboration has been built into the design from their very conception, in other cases the need for collaboration will arise in the same informal fashion as has arisen in the past. At the same time, many modern influences in medicine (e.g. evidence-based practice) have led to the definition of 'protocols' and 'care pathways' which may readily be recognised as workflows, thus providing a context for some collaborations. None the less, there is a good deal of scope for knowledge and technology transfer from heavily data-driven branches of e-science, where collaborative workflow engines have begun to be established.

Table 3. Health Research challenges for a collaboration grid

Research challenge name	Description of the health research challenges
RCLG1	Migration of e-science workflow engines to biomedical research to encompass end-to-end processes, e.g. stages on the road from drug discovery to clinical trial.
RCLG2	Natural mapping of healthcare/medical protocols to workflows for remote collaboration, education or quality control.
RCLG3	Certification of medical workflows, complying with relevant legal and ethical obligations, to ensure they are reliable, validated and updated when required.
RCLG4	Natural mapping of public health distributed decision support to facilitate coordinated action.
RCLG5	Natural mapping of guidelines, protocols and integrated care pathways to validate practice against constantly updated evidence base.
RCLG6	*Ad hoc* integration of heterogeneous sources of information where no prior coordination has been provided. Integration of different levels or modalities of medical data towards multidisciplinary diagnosis and treatment planning.
RCLG7	Workflow repositories to retain and maintain defined workflows and to enhance reuse, repurposing and recycling. Retain workflow histories and outcomes.
RCLG8	Support for persistent collaborations, esp. in relation to rights management and participant privileges.
RCLG9	Integration and management of workflows with implications in different domains, e.g. conflict between medical and ethical calls.
RCLG10	A forum for the discussion of health/medical workflows, including provenance data, and a broader means of discussion and communication between collaborators.

Issues of collaboration arise in the context of diagnosis by different specialists, second opinion, treatment and surgery planning. Examples would include pipelining second reading or second opinion in breast screening; bringing in additional expertise if appropriate – e.g. staging a cancer; or quality control of the consistency of histopathology findings, by analysis of reports and checking them against guidelines. Monitoring in the public health domain may be concerned with MRSA-type epidemics or with avian flu or heat wave emergencies. All these call for a different kind of joint action, but all may benefit from decision support. More sophisticated epidemiology may be possible through analysis of associated data, as in the suggestion that avian flu passes more readily among genetically related individuals than among others despite close contact [51]. These suggested requirements would be satisfied through a combination of knowledge management and workflow management tools, linking the two where necessary.

In another dimension of collaboration, the promise of modern biomedicine to relate genomic data to disease phenotypes, is being explored in such projects as Health-e-Child (HeC) and ACGT [52]. In HeC, there is a need to bring together information not only from different levels but also from different modalities, such as genome and imaging data. Thus collaboration here will also mean ability to coordinate different tools and modalities as well as integration of knowledge and data.

Finally, a further development is possible in the context of this discussion, to coordinate 'publication' of services and certification/licence issues.

5.1.4. Health Research Challenges for Knowledge Grids

Table 4 provides a list of research challenges for knowledge grids (RCKG). These challenges refer repeatedly to data integration and knowledge management.

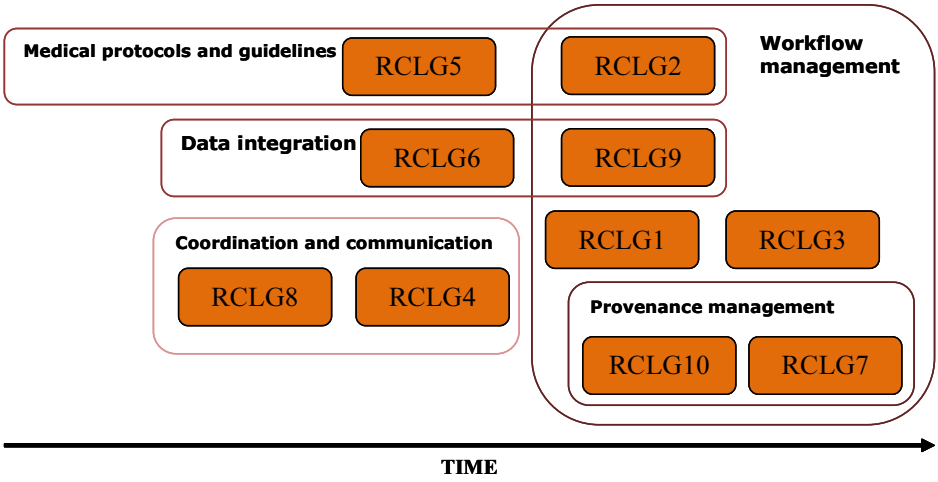


Figure 4. Dependencies for collaboration grid research challenges.

Table 4. Research challenges for a knowledge grid

Research challenge name	Description of the health research challenges
RCKG1	Knowledge-driven grid catalogues and integration based on the metadata.
RCKG2	Develop standards and models for exposing web services (semantics), scientific services, and the properties of data sources, data sets, scientific objects, and data elements
RCKG3	Design standards for and build an expert tool (ontology/schema/rules negotiator) for exposing the properties of local sources in a federated environment
RCKG4	Develop enhanced knowledge representation models and data exchange standards for complex systems, presently largely inconsistent or incomplete, looking for synergies with other initiatives
RCKG5	Develop new, domain-specific ontologies, built on established theoretical foundations and taking into account current initiatives, existing standard data representation models, and reference ontologies
RCKG6	Design standards for and build an expert tool (services/data negotiator) to guide users through the complexities of the data, data models, simulation and modelling tools, etc.
RCKG7	Develop advanced text mining tools for capturing implicit information about complex objects, relationships and processes, as described in patents and literature, beyond and above simple pair-wise relationships between entities

Many of these challenges include the definition of standards and ontologies (RCKG2, RCKG3, RCKG4, RCKG5, RCKG6). Some challenges are directly related to the grid technology itself (RCKG1, RCKG2, RCKG3) while others are more relevant to the research area (RCKG4, RCKG5, RCKG6, RCKG7) and therefore not specific to the grid technology. In that case, it seems the healthgrid should benefit from the knowledge management services once they have been developed by the research community.

The key words we will keep to characterise the research challenges for a knowledge grid are data integration tools and standards as well as knowledge management tools and standards. In addition, we will use the concept of domain specific knowledge

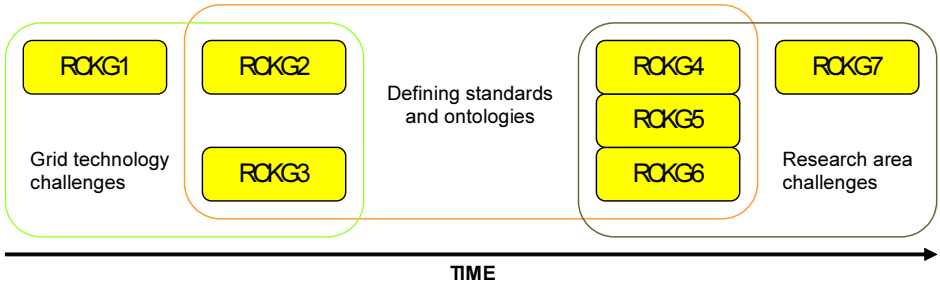


Figure 5. Dependencies for knowledge grid research challenges.

Table 5. Deployment milestones

Deployment Milestone name	Description of the milestone
MD1	Need for successful pilot applications on epidemiology and VPH that will demonstrate the benefits of the technology to foster adoption of grids in the community and to identify limitations of existing infrastructures.
MD2	Need for epidemiology data sources adapted to grid models.and grid-enabled gateways to epidemiological data using medical informatics-related connectors, such as HL7, DICOM, ENV13606, etc.
MD3	Build a core reference database of validated experimental and clinical research data extracted from the literature
MD4	Creation of disease-specific European Imaging Networks for establishment of standards, validation of imaging biomarkers and development of regional centres of excellence.

management tools and ontologies to characterise the developments which are not specific to grids but are needed to enable a knowledge grid.

The dependency diagram for these milestones, grouped by areas, can be seen in Fig. 5.

While work has been done towards many of these milestones, they remain significant challenges due to incomplete implementations and immature standards. Many of these challenges are as much in the Artificial Intelligence domain as in grid computing, with issues ranging from 'knowledge-driven' resource and service management to ontologies and meta-ontologies for medical knowledge.

5.1.5. Deployment Milestones

Table 5 provides a list of deployment actions which were recommended by the research communities. These actions are perceived as milestones on the road to healthgrid adoption as their success will pave the way to the adoption of the technology.

Some actions are more geared towards computing grids (MD3) some are related to data grids (MD1, MD4) while others require from the beginning knowledge management (MD2, MD5).

These actions could be started on the existing grid infrastructures in view of the present state of the art of the grid technology. However, the quality of the services as well as their portfolio is expected to improve with the evolution of the technology.

5.2. Proposed Roadmaps

In this section, we are going to present a technical roadmap for the adoption of the grid technology for healthcare. In the previous section, for three families of grids, computing, data and knowledge grids, we have identified a number of research challenges which have been characterised by a few key words.

Computing grids:

- user friendliness
- interoperability of infrastructures
- quality of service
- on demand access.

Data grids:

- improved distributed data management
- quality of service
- distributed data models.

Knowledge grids:

- data integration tools and standards
- knowledge management tools and standards
- domain specific knowledge management tools and ontologies.

Extending the model of Fig. 1, Fig. 6 represents how research challenges address different layers of services from core infrastructure to applications. The following comments can me made from the picture:

- Interoperability as well as improved distributed data management must be core functionalities of the infrastructure.
- Quality of service is required from both core and healthgrid services for successful healthcare / biomedical applications.
- Healthgrid services should be accessible on demand, in a user friendly way. Distributed data models need to be provided as well.
- Data Integration Tools and Standards are healthgrid services which stand at the interface between data and knowledge grids.
- Knowledge Management Tools and Standards require the availability of proper job and data management tools. They stand at the interface between generic healthgrid services and the application specific developments.
- Domain Specific Knowledge Management and Ontologies are under the responsibility of the research communities. Their interface to the knowledge grid is achieved using the Knowledge Management Tools and Standards.

On the basis of this analysis, we have represented in Fig. 7 the research challenges according to their complexity and an estimated time when they should be overcome. The figure inspired from The Innovative Medicine Case Study,[2] also indicates the level of adoption by the research communities. As can be seen clearly from the picture, we identify several distinct roadmaps:

[2] SHARE deliverable D5.2a.

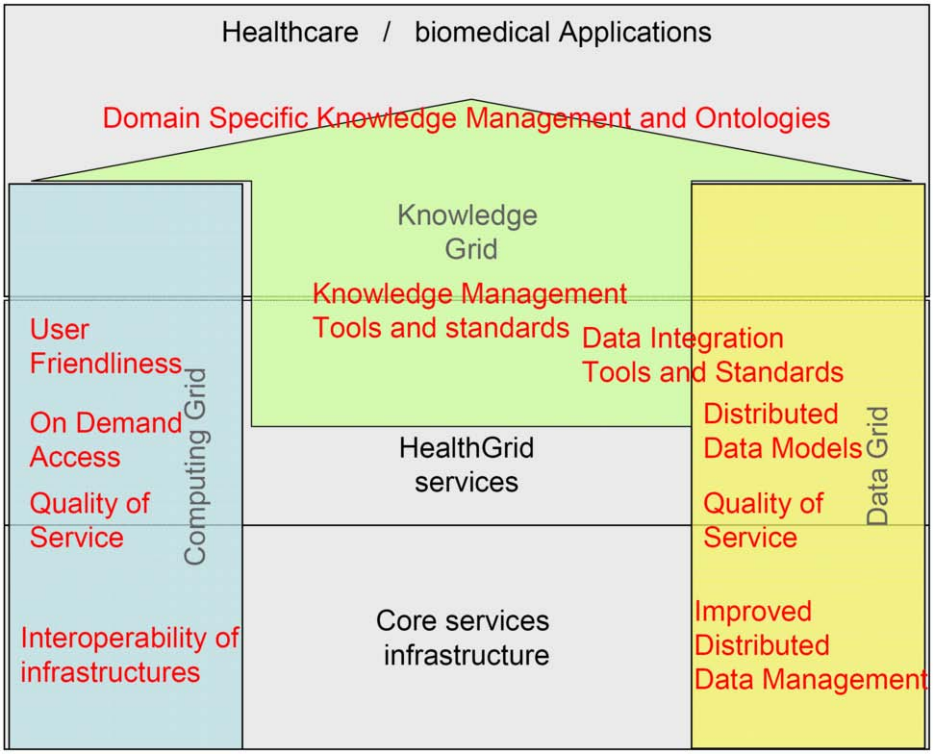


Figure 6. Representation of research challenges and healthgrid layers of services.

- research and development for computing grids should allow offering the quality of services needed for biomedical research and healthcare at a 5-year horizon.
- data grids are expected to reach maturity at a 10-year horizon as the core technology is not yet mature.
- collaboration grids are achievable with different levels of sophistication at different stages.
- knowledge grids depend on the quality of services for distributed data integration and the capacity of the research communities to agree on standards and ontologies. As a consequence, their maturity is not expected before 15 years.

This model arises in the field of innovation studies, and distinguishes between: *visionaries*, pioneers both in research and applications; *early adopters,* who recognise the potential for rapid benefits and take up the technology quickly, often introducing further innovation; *early majority* who incur relatively little risk in adopting the technology; and *late majority,* those who take virtually no risk and finally adopt a technology because there is no other option.

The diagram depicts priorities: even for early adopters, infrastructure interoperability and distributed data management are already necessary; on demand access, usability and quality of service are at the first point of inflection, before rapid expansion, while with sophisticated AI tools in the later stages, a second inflection occurs and the technologies become routinely accepted.

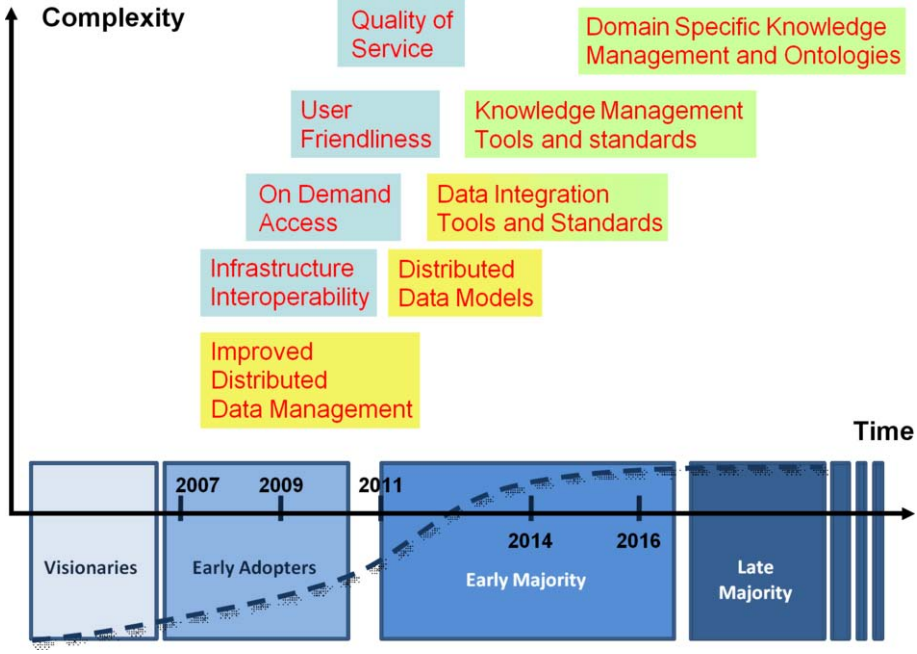
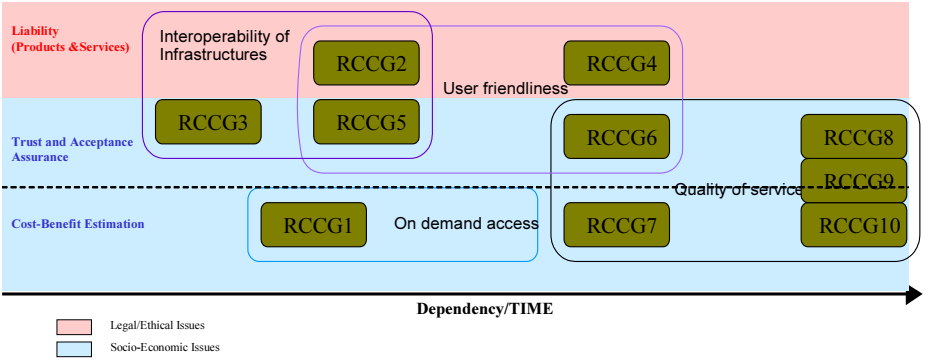


Figure 7. Research challenges as a function of time and complexity.

5.3. Mapping ELSE Requirements

5.3.1. Computing Grids Phase

Research challenges for Computing grids



5.3.1.1. Liability Issues

The current state of EU legislation does not cover liability issues that are specific to healthgrids. The following tasks could help minimise liability concerns for healthgrid usage.

- Analysis and prediction of risks and possible damage to patient health and privacy should begin from the outset.
- Outlining examples of liability concerns specific to the use of healthgrids in order to encourage the introduction of new legislation and policies.
- Good testing strategies for services and products, including testing for product safety.
- Good quality assurance for services and products.
- The use of standard techniques for detecting bugs and faults while dealing with infrastructure interoperability.

5.3.1.2. Trust and Acceptance

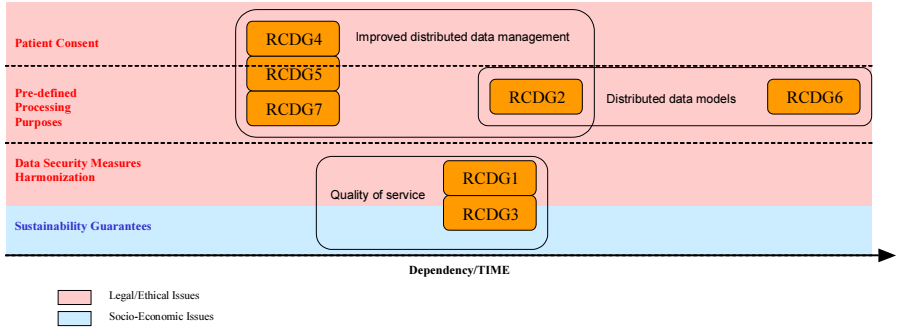
- Pilot projects and prototype applications to demonstrate the use of grid services in clinical and research workflows
- The use of trust based technologies to increase the trustworthiness of grid infrastructures
- Providing feedback and documentation to provide users with clear answers to any security concerns.

5.3.1.3. Cost and Benefits Estimation

- Ex-ante analyses over time, based on initial pilot experiences. These have to focus on ensuring acceptance, technical and regulatory certainty, and sufficient private incentives in the steps to follow
- Analysis to estimate potential net benefits (i.e. expected benefits less expected costs over time), accounting for different risks and for optimism bias in estimations. Such studies will facilitate access to initial funding, but can also be beneficial in the necessary dissemination work with the health sector.

5.3.2. Data Grids Phase

Research challenges for Data grids



5.3.2.1. Data Protection

- Ensuring the use of standard means of data security within the different data management systems of participating infrastructures (encryption, anonymisation, pseudonymisation, access control, etc.).

- Use of data quality assurance mechanisms.
- Adopting formal ways to audit the regulatory compliance of operational level controls.
- Automation of the collection of patient consent, and ways to allow patients to opt in, opt out and withdraw consent.
- The use of evolving privacy enhancing technologies.

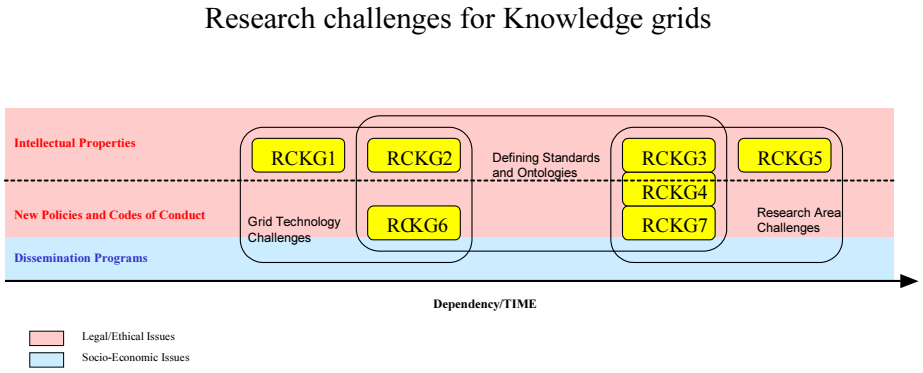
5.3.2.2. Sustainability Guarantees

- The development and deployment of data grids will benefit from more focused prospective assessments of the socio-economic impact in order to identify existing and potential barriers.
- Convincing business cases ensuring sustainability.
- An organisational milestone can be defined here in the move from technology/science towards service provision. By that stage, a notable amount of legal and regulatory certainty has to be achieved, so that private incentives can be assessed and adjusted (including via government intervention) if necessary.

5.3.2.3. Education and New Skills Requirements

- Training and educational programs to increase users' confidence in the use of healthgrid products and services.
- Adequate documentation and guidance must be available, and where grid infrastructures are distributed amongst geographically remote sites there must be sufficient communications methods such as video conferencing to ensure problems and concerns raised during deployment are dealt with quickly and efficiently.
- Investing in technical staff within hospital and research centres to provide help with technical problems.

5.3.3. Knowledge Grids Phase



5.3.3.1. Intellectual Property

- There is a contradiction between intellectual property rights and the needs of grid technology, which will require that access to databases, knowledge and software is free of rights. Contract law and agreements could be an option to regulate the IP issues related to knowledge integration, ontologies and software reuse.

5.3.3.2. Governance and Delegation

- The working document on the processing of personal data relating to health in electronic health records (EHR) recommends that in the case of health care systems that adopt a decentralised data storage model, it could be necessary to appoint one central body to be responsible for steering and monitoring the whole system and also for ensuring the operation of the system is compatible with data protection. It would also be useful if data subjects could address their data protection queries to a central body instead of having to search and identify the relevant controller among many. The architecture of a healthgrid system is similar but even more complex than a distributed system. The idea of one data controller might be preferable but more challenging. A discussion and analysis highlighting the main issues and benefits surrounding the idea of a unique data controller for data stored within a healthgrid domain needs to be produced. Linking this to the technological component of the roadmap, this could impact on the process of granting permission to access the data.

5.3.3.3. Policies and Codes of Conduct

- Discussions should take place between different healthgrid stakeholders to decide on the importance and benefit of applying for new legislation to address healthgrid related legal and ethical issues.
- Once a decision is reached, a framework will need to be distributed to legal bodies showing why healthgrid services and products should be considered different from other marketed products. It could also present scenarios showing that the current legislation does not ideally cover these issues.
- Evolving technologies for the automation and enforcement of policies at different infrastructure layers should be explored.

5.3.3.4. Dissemination and Publicity

- Dissemination and publicity programmes need to precede the deployment of knowledge grids. This includes workshops, conferences, and magazines to attract the user community and build awareness of healthgrid facilities
- Demonstrate the effectiveness of grid applications in providing healthcare and research services while preserving the users' autonomy.

5.3.3.5. Ethical Control and Auditing

- In the UK, every health organisation is now required to have a privacy and data protection officer, the so-called "Caldicott guardian". The establishment of similar roles throughout Europe would be a major step towards harmonisation of ethical practice and compliance in the member states. This may be supplemented by the creation of a Europe-wide ethical body composed of these European Caldicott guardians, although there is a question about how this would relate to the article 29 working party. Operating at a healthgrid level, they would be able to judge matters in a European context. This would in a sense provide a useful bottom-up approach to confidentiality and privacy protection across healthgrids, as opposed to top-down European directives
- Before the deployment of Data grids should start, the requirements for ethical oversight and monitoring should be determined, and the automation of oversight facilities should be explored.

5.3.4. Benefiting from the Technology

Despite the best efforts of policy makers to structure and control the use of personal data (in this case patient data), incidents of identity theft, identity base fraud, and the sale and misuse of data are still occurring, as highlighted by recent stories in the media. This may in part be due to a lack of enforcement of high-level legal obligations concerning personal data, and also as a result of the variability of privacy laws due to cultural and national considerations.

The recent push to enforce legal rules within enterprise infrastructures and business processes has initiated several EU and international projects, each aiming to help enterprises, organisations and governments to benefit from the use and exchange of personal data without compromising individual privacy. As a result, many privacy-enhancing technologies have been developed which have proven efficiency across many domains including e-health. The possibility of deploying similar technologies on a grid infrastructure should therefore be investigated.

5.3.4.1. Privacy-Enhancing Technologies (PETs)

The approach to privacy typically taken by more advanced systems involves enforcing privacy policies at the application level; the filtering of sensitive data before providing the query result to the user. But this could still reveal enough information for an intelligent person to identify individuals. For better protection of the data, access control policies have to be enforced at the data level. Traditional databases provide access control at the table level and use the view mechanism to restrict access to certain columns or rows of the table, but this is still inadequate. Hippocratic databases [50] provide a more advanced "limited disclosure" approach. They permit enforcement at a very fine level of granularity. Privacy policies could be enforced at the level of an individual cell in a relational table. Hippocratic databases also allow privacy policies to be stored and managed in the database as metadata.

Sticky policies [51] have emerged as one approach to enhance privacy preservation in distributed computer systems. The underlying notion behind *sticky policy enforcement* is that the policy applicable to a piece of data travels with it and is enforceable at each point it is used. Recent work done by the IBM Almaden Research Laboratory has improved this approach, making it adequate for the needs of a healthcare environment.

They have added new functionality to handle data disclosure to a party with well-defined constraints that allows data to be released to less privileged parties without requiring the originator's involvement. A crucial task prior to the disclosure was identifying the applicable privacy policy constraints for the document(s) to be shared and sticking them together, forming a single entity for transfer. This avoids the potential pitfall of having to contact a (potentially) large number of third parties before making a decision to disclose a specific piece of information.

The PRIMA (PRIvacy Management Architecture [52]) System was developed by the IBM Almaden research lab in order to exploit *policy refinement* techniques to gradually and seamlessly embed privacy controls into clinical workflows based on the actual practices of the organisation in order to improve the coverage of the privacy policy. PRIMA attempts to improve policy coverage by gradually embedding new policy statements, which were discovered through the process of policy refinement, into the clinical system. Stakeholders define the privacy policies, which are embedded in *privacy controls* that are integrated into the clinical environment. One of these privacy controls is an auditing function that automatically generates entries for the system's audit logs. These logs are either periodically replicated or PRIMA-enabled by the con-

struction of a consistent, consolidated view of them. In the simplest case, there is just one log. At regular intervals or at the request of stakeholders, the *policy refinement* component extracts input from the *audit management* component and the *privacy policy definition* component and outputs a list of definitions, if any exist, that should be included in the policy definitions.

Enterprise Privacy Authorisation Language (EPAL) [53] was designed to enable the translation of privacy policies into an XML based computer language. The resulting coded translation of human policy into information technology policy allows complex descriptions of the internal data handling practices needed for enforcing the privacy policy. The expressiveness of EPAL has been tested against a set of "real world" scenarios such as of the Ontario Freedom of Information and Protection of Privacy Act (FIPPA). This has demonstrated the effectiveness of EPAL in

- Linking access control to natural text policies
- Creating precise, fine grained description of the policy
- Enabling complex, context driven conditions on policy rules
- Creating portable, reusable policies
- Allowing for sector/legislation specific policy vocabularies
- Enabling policy negotiation.

5.3.4.2. Trust-Based Technologies

To qualify as a trustworthy system [54], the healthgrid security infrastructure needs to adopt technologies that are able to fulfil the following needs of users and resource providers. Before sending the job request, the user needs to:

- Know whether the resource provider host in the resource provider domain (to be visited) is "trustworthy" in terms of faithfully executing the user code and completing the task.
- Know whether the resource provider host(s) will have enough trust in the user to cooperate with them (i.e. a *code trust* question involving *trust symmetry* problem). In many cases such a trust relationship is often implicitly assumed.
- Ensure that the resource provider host(s) will not tamper with the user code and/or computation result.

Before running the job request on the resource provider node, there can be two *code trust* questions that the resource provider node should ask:

- Is the job requesting user trusted to produce benevolent and competent code that will not harm the grid?
- Has the user program been tampered with before it is allocated?

After completion of the job result:

- Both user and resource provider(s) need to update their relevant trust relationship knowledge.
- The user needs to check the integrity of the completed job or result to update its *execution trust* with the resource provider(s). Resource providers need to update the *code trust* for the user.

PETs have shown they can be efficient when deployed across a variety of domains. However, each technology typically only deals with a stand-alone privacy or security

issue. In order to optimise patient privacy protection, many PETs will need to be integrated into a single privacy framework.

The PRIME project,[3] an EC FP6 project, aims to reconcile privacy and accountability of users' electronic interactions in Europe. The project addresses these goals by providing an architecture integrating several privacy enhancing technologies and emerging systems that include human-computer interfaces, ontologies, authorization and cryptology, anonymous communication, privacy-enhancing identity management architecture, and assurance methods. The PRIME project also recognised the need for solutions to be compatibe with the existing legal framework in order for them to have real world relevance. Therefore legal requirements were considered from a very early stage, and the PRIME solution integrated legal rules to be more efficient and to form a "privacy-protecting framework that has a real impact on business practices".

6. Concrete Recommendations

6.1. Technical Recommendations

It is important that technical research and development be conducted in close collaboration with user communities. At certain stages it must be driven and validated by user groups, although there is always scope for innovators to introduce unforeseen possibilities to users. The research communities involved in the definition of the roadmap expressed their interest and support for the deployment of prototypes and test cases on existing grid infrastructures. We recommend that these infrastructures and tools continue to be made available to applications requiring computing services and data management.

Indeed, some projects are already using the DEISA and EGEE infrastructures for scientific production in the fields of epidemiology, medical imaging and drug discovery. However, these initiatives come from pioneers and are not sufficient to achieve a wider adoption in these research communities. We recommend that:

- More attention be paid to such initiatives so that they may influence the evolution of the technology to make it better fit the needs of the community;
- Two projects within the framework of the EuroPhysiome initiative be identified that could directly benefit from the computing and data management resources of the EGEE and DEISA infrastructures; these should be deployed in parallel on the two infrastructures in order to investigate interoperability issues and identify bottlenecks.

In terms of encouraging biomedical applications to fully exploit grids, we recommend:

- Linking certain advanced health domains to an e-science infrastructure;
- The adaptation of epidemiology data sources to grid models and grid-enabled gateways to epidemiological data, using medical informatics-related connectors such as HL7, DICOM, ENV13606, or similar.

[3] See https://www.prime-project.eu/.

In the same spirit, in order to foster the uptake of grids in the biomedical research and healthcare communities, we recommend:

- The release of open-source components for medical data interfacing;
- Building a core reference database of validated experimental and clinical research data extracted from the literature in innovative medicine and to explore whether a grid infrastructure could support this activity;
- The creation of disease-specific European imaging networks towards the establishment of standards, validation of imaging biomarkers, development of regional centres of excellence in innovative medicine and exploration of grid infrastructures to support such activity.

We recognise that there are a number of concerns (for example: security and standards) in which problems exist irrespective of the use of grids. It is important to understand the nature of these problems and the extent to which the use of grids complicates them. Results could be concrete implementation recommendations (for example: security improvement) and a suggested list of health applications requiring security which may be able to be deployed on a grid. In the field of standards, we believe that the HealthGrid initiative provides the right framework to coordinate the development of the different standards in collaboration with the OGF and the various medical informatics standardisation bodies. We recommend:

- The active pursuit of standards for the sharing of medical images and electronic health records on the grid within the already existing medical informatics standardisation bodies;
- The active pursuit of ontology matching and development for healthgrids;

We believe that technology transfer between EC projects should receive more prominent and active encouragement. In particular, we recommend:

- The commission implements collaboration measures in the funding mechanism for projects;
- Targeted capacity building so that projects may access grid resources on demand, without previous agreement or request; European grid infrastructures should be freely accessible to European projects;
- Porting of one or two biomedical grid applications, already successfully deployed on grid infrastructures, to e-science environments using OGSA-compliant grid toolkits.

Finally, to return to a frequent theme in this analysis, we recommend:

- The encouragement of cross community interaction, in order to build meaningful dialogue between grid developers and health researchers.

6.2. Legal Recommendations

Liability in a healthgrid System Using grids blurs the liability issues in terms of medical practice. A stepwise approach should therefore be taken to develop the liability framework, distributing legal responsibility appropriately across healthgrid users. Such an approach would help to favour the reliance on the system while providing legal certainty for all stakeholders, including patients. Moreover, the European Commission should consider supporting the adoption of EU level guidelines that would identify the

various parties involved in delivering healthgrid services and annex services and establish the various liabilities that each party must accept. Such guidelines should be widely disseminated in order to develop users' confidence in the use of healthgrids in general. In particular it should be investigated whether specific guidelines on those specific services could be drafted under the provisions for a code of conduct established in directive 2000/31 on eCommerce [46].

Product safety As mentioned in D4.2, in the framework of the European level legislation applicable to product safety, national authorities have been established to monitor product safety and to take appropriate measures to protect consumers. Under these circumstances, an information system has been put in place that imposes collaboration between distributors, producers and the national authorities but also between member states and the European Commission (RAPEX) [55].

At present, this system is not used at all for products used in the composition of grid systems. The European Commission should thus adopt policy tools encouraging the use of the RAPEX system for such products.

Healthgrid as a Medical Device As outlined in the introduction to this document, the law on medical devices is very unclear with respect to healthgrids. While it may be argued that a healthgrid could fall within the ambit of the current medical devices directive [47] in that it is a software tool that impacts on a medical act, the whole construction of the directive is based upon physical goods (which might have a software component) that are placed on the market for purchase or lease. In this situation, many of the currently available monitoring devices are covered only by general product liability, but not by specific liability provision.

In this framework, special guidelines should be issued in order to clarify the application of medical devices legislation to specific tools used in healthgrids.

Patient Consent In February 2007, the European working party on data protection, established under article 29 of the directive issued a working paper looking at the applicability of data protection legislation to Electronic Health Record (EHR) systems [48]. In its report, the working party noted in particular the limitation of the use of consent to permit the processing of heath data. The working party notes that if processing health data in an EHR system is the primary way of processing health data in a given health system, then a patient's care may be compromised if he or she opts-out of such a system by not giving his or her consent to the creation of an EHR. Accordingly, consent should not be used as it cannot be said to be truly and freely given.

The remaining provisions setting aside the general prohibition on article 8 of the directive 95/46/CE [49] can also be said to pose some problems – notably the idea that a patient ought to know the full finality of the use of data before his or her data may reasonably be used. But, as noted by the data protection working party there are some problems in using consent as a valid basis for processing data in eHealth applications. Indeed, if the creation of, for example, electronic medical records is a necessary and unavoidable consequence of the medical situation, withholding consent may be to the patient's detriment.

Specified and Explicit Purposes According to the data protection directive [49], data may only be collected for specified and explicit purposes. If healthgrids can be used for risk detection, disease monitoring and preventive care, legal guidelines should be established that clarify the circumstances in which professionals can make further use of personal data related to health in the interests of public health. Such guidelines should allow for secondary uses even where such uses could not have been foreseen at the time of data collection.

Technical and Organisational Security Measures Efforts should be made to harmonise national standards on the technical and organisational measures of data security. While the data protection directive calls for such standards to be adopted, little has been done at a regulatory level to harmonise guidelines across the EU.

Intellectual Property Rights It might be desirable for the commission to develop guidelines for the use of open licensing and open standards, which could address the tension between the intellectual property rights of developers and the needs of the grid technology. Such an open standards software approach could then be a solution to help the development and implementation of healthgrids.

On the other hand, the use of healthgrids in the drug discovery sector raises the issue of the ownership of both methods used to discover the medicines and the results achieved. Indeed, all the grid nodes that contribute resources to compute the docking probabilities could claim some ownership of the results and the designers of the software used in the process would certainly be in position to claim ownership of the method. In this context, one may ask whether it is important to know, say, which grid node was the one to identify a particular candidate molecule.

In this context, it is of essential interest, notably in patents, to determine guidelines that would determine, in case of collaboration in the research, what every actor is entitled to according to his contribution to the system.

Privacy Policies and Codes of Conduct As suggested above, a directive or code of conduct on privacy and health information infrastructure should be developed within the context of directive 95/46/EC and could take the form of either a dedicated directive or could be an EU-level code of conduct to be approved by the European working party on data protection set up under article 29 of the directive. This could help to solve the problem of data processing legitimacy. In particular, it could provide possible bases of legitimacy other than the data subject's consent. It could also provide the following solutions:

- Appropriate safeguards to allow for the further processing of personal data (and especially of medical data) for substantial public interests (without requiring the data subject's consent) like scientific research. An example of appropriate safeguard would be a first coding by the initial data controller and a second coding by a trusted third party gathering all the data from the data controllers before sending them to the researchers.
- Appropriate safeguards to allow keeping the data for longer periods for scientific use; terms under which identification numbers or other identifiers may be used; terms under which (coded) personal data may be transferred to third countries for scientific research.

6.3. Socio-Economic Recommendations

Trust and Acceptance Trust is a very important element in any interaction between the different members of a society. In the market context, trust is crucial for successful business to business collaborations. Similarly, in a healthgrid domain a good collaboration will not be achieved unless a trust relationship exists between the different users and stakeholders. Pilot projects and prototype applications, which are an inherent part of the technology roadmap, need to be future oriented in the sense that the ultimate routine operation users have to be persuaded both of their value and their applicability, i.e. their ability to fit into real clinical or research workflows. This has to be taken seri-

ously from the very beginning, even in proof-of-technology demonstrators: the goal should always be to give users, especially clinicians, tools that they would consider using with patients in real healthcare situations. Trust and acceptance can be greatly enhanced by the establishment of appropriate ethics committee structures to advise on the observance of ethical principles.

Estimation of Costs and Benefits Ex-ante analyses over time, based on initial pilot experience, have to focus on ensuring acceptance, technical and regulatory certainty, and sufficient private incentives in the steps to follow. An inherent part of such assessments should be to estimate potential net benefits (i.e. expected benefits less expected costs over time), accounting for different risks and for optimism bias in estimations. Such studies will facilitate access to initial funding, but can also be beneficial in the necessary dissemination work among the health sector.

Sustainability Guarantees Work towards achieving the next milestone in complexity – data grids – will benefit from more focused prospective assessments of socio-economic impact in order to a) identify already existing, as well as potential barriers, and b) build convincing business cases ensuring sustainability. The analysis of alternative resource allocation options from a societal perspective, but also on organisational level, becomes necessary.

An organisational milestone can be defined here in the move from technology science towards service provision. By that stage, a notable amount of legal and regulatory certainty has to be achieved, so that private incentives can be assessed and adjusted (including via government intervention) if necessary.

Cross-Organisational Interoperability The effective deployment of knowledge grids will crucially depend on collaboration between institutions, meaning more than "simple" access to each others' data and computing resources. This collaboration requires the utilisation of human resources and in some cases a significant strategic re-orientation and re-organisation of working processes and even management structures. As the health sector, including clinical research and public health, is (and should be) highly regulated, policy makers on regional, national, and EU level should review the existing regulatory framework against the requirements arising from the exploitation of knowledge grids. Particular attention should be given to flexibility of government regulated budgets and reimbursement schemes. The latter should encourage cross-organisational collaboration, including such beyond national borders, by means of using knowledge grids.

7. References

[1] Information on HealthGrid initiative available at http://www.healthgrid.org.
[2] V. Breton, K. Dean and T. Solomonides (eds.), The HealthGrid White Paper, Proceedings of Third HealthGrid Conference, IOS Press, Vol. 112 (2005) pp. 249–321; preliminary version CISCO/ HealthGrid paper http://www.healthgrid.org.
[3] Information on Share project available at http://www.eu-share.org.
[4] Action plan for a European e-Health Area, COM(2004) 356, European Commission, http://europa.eu.int/information_society/doc/qualif/health/COM_2004_0356_F_EN_ACTE.pdf.
[5] V. Breton, A. E. Solomonides & R.H. McClatchey. A Perspective on the HealthGrid Initiative. Second International Workshop on Biomedical Computations on the Grid, CCGrid 2004, Chicago, USA, April 2004.
[6] A. Sousa Pereira, V. Maojo, F. Martin-Sanchez, A. Babic & S. Goes. The INFOGENMED Project. ICBME 2002, Singapore, 2002.

[7] F. Martin-Sanchez et al. *Integrating Genomics into Health Information Systems*. Methods of Information in Medicine 41: pp. 25–30, 2002.
[8] BIOINFOMED Project. *Synergy between Medical Informatics and Bioinformatics: Facilitating Genomic Medicine for Future Healthcare*, BIOINFOMED Study Report White Paper, 2003.
[9] Present status and future prospects of Grid contribution to health care. *Virtual Medical Worlds*, Virtual Magazine on Telemedicine and High Performance Computing and Networking. http://www.hoise.com/vmw/04/articles/vmw/LV-VM-04-04-24.html.
[10] Dr Wolfgang Leister, Dr Asmund Skomedal, "Grid Systems in Health Care", Norsk Regnesentral Norwegian Computing Centre.
[11] Ignacio Blanquer & Vicente Hernández. *The Grid as a Healthcare Provision Tool*. Methods of Information in Medicine 44: pp. 144–148, 2005.
[12] SHARE Public Deliverable D4.1 See http://eu-share.org/fileadmin/templates/Document/SHARE-D4.1.pdf.
[13] Treaty of the European Union Art. 152 provides that matters of health services organisation are subject to the rule of subsidiarity and limits the role of the EU to supporting and co-ordinating the activities of the Member States.
[14] "Council Conclusions on Common values and principles in European Union Health Systems", Document (2006/C 146/01), Official Journal of the European Union on 22 June 2006, pp. 1–5.
[15] Christine Huanga, Laura A. Noirotb, Kevin M. Hearda, et al. Implementation of Virtual Medical Record Object Model for a Standards-Based Clinical Decision Support Rule Engine. Proceedings of AMIA Symposium (2006).

For all SHARE documents see http://eu-share.org/about-share/deliverables-and-documents.html.

8. Terminology

8.1. Abbreviations

DICOM	The Digital Imaging and Communications in Medicine Standard
EC	European Commission
EHR	Electronic Health Record
ELSE	Ethical, Legal and Socio-Economic
EPAL	Enterprise Privacy Authorization Language
EU	European Union
HL7	The Health Level 7 Standard
HPC	High Performance Computing
IPR	Intellectual Property Rights
OGSA	Open Grid Services Architecture
QoS	Quality of Service
SHARE	Supporting and structuring Healthgrid Activities and Research in Europe
SOKU	Service Oriented Knowledge Utility
VPH	Virtual Physiological Human
W3C	World Wide Web Consortium
WPx	Work Package x
WP3	SHARE Technology and Security Activity
WP4	SHARE Health Policy, Legal, Social and Economics Activity
WP5	SHARE Applications Activity
WP6	SHARE Roadmap Synthesis and Validation Activity

8.2. Definitions

The following definitions are useful for understanding the document content.

- **Authentication**: Verifying and confirming the identity of a grid user.
- **Authorisation**: Restricting access to resources based on what a user has been granted access to.

- **Data**: Any and all complex data entities from observations, experiments, simulations, models, and higher order assemblies, along with the associated documentation needed to describe and interpret them.
- **Data controller**: The person or organisation responsible for the manner in which any personal data is processed.
- **Data mining**: Automatically searching large volumes of data for patterns or associations.
- **Data model**: A model that describes in an abstract way how data is represented in an information system. A data model can be a part of ontology, which is a description of how data is represented in an entire domain.
- **Data processor**: Any person who processes data on behalf of a data controller.
- **Data subject**: An individual who is the subject of personal data.
- **Grid**: A fully distributed, dynamically reconfigurable, scalable and autonomous infrastructure to provide location independent, pervasive, reliable, secure and efficient access to a coordinated set of services encapsulating and virtualising resources.
- **Informed consent**: A legal term referring to a situation where a person can be said to have given their consent based upon an appreciation and understanding of the facts and implications of an action.
- **Metadata**: May be regarded as a subset of data, and are data about data. Metadata summarise data content, context, structure, inter-relationships, and provenance (information on history and origins). They add relevance and purpose to data, and enable the identification of similar data in different data collections.
- **Middleware**: A software stack composed of security, resource management, data access, accounting, and other services required for applications, users, and resource providers to operate effectively in a grid environment.
- **Ontology**: The systematic description of a given phenomenon, which often includes a controlled vocabulary and relationships, captures nuances in meaning and enables knowledge sharing and reuse. Typically, ontology defines data entities, data attributes, relations and possible functions and operations.
- **Processing**: Obtaining, recording or holding the data, or carrying out any operation on the data, including organising, adapting or altering it. Retrieval, consultation or use of the data, disclosure of the data, and alignment, combination, blocking, erasure or destruction of the data are all legally classed as processing.
- **Roadmapping**: An extended look at the future of a chosen field of inquiry, leading to an outline or map of how and by what means to achieve certain goals.
- **SOAP**: A protocol for exchanging XML messages over a network. It defines the structure of the XML messages (the SOAP envelope), and a framework that defines how these messages should be processed by software.
- **The Article 29 Data Protection Working Party**: A working party established by article 29 of directive 95/46/EC. It is the independent EU advisory body on data protection and privacy. Its tasks are laid down in article 30 of directive 95/46/EC and in article 14 of directive 97/66/EC.
- **Virtual Organisation**: A group of grid users with similar interests and requirements working collaboratively and/or sharing resources regardless of

geographical location.

- **Web Service**: A software system designed to allow inter-computer interaction over a network to perform a task. Other computers interact with a web service, in a manner prescribed by its interface, using messages which are enclosed in a SOAP envelope and are often conveyed by HTTP. Software applications can use web services to exchange data over a network.
- **Workflow**: A set of components and relations between them, used to define a complex process from simple building blocks. Relations may be in the form of data links which allow the output of one component to be used as the input of another, or control links which state some conditions on the execution of a component.
- **XML**: An annotation technology used to describe structured data within a document using mark-ups and tags, similar to HTML. The main difference between the two is that the elements in XML can be given a definition depending on their usage which may be semantic rather than presentational. XML is a text format and can be read easily either by humans or machines.
- **XML Schema**: A definition of the structure of an XML document. A schema contains a set of rules that dictate how an XML document must look like in order to be an instance of this schema. The relationship between a schema and an XML document implementing it can be compared with a class definition and an instance in object-oriented programming.